

Film classification based on low-level visual effect features

Hui-Yu Huang

National Formosa University
Department of Computer Science and Information Engineering
64, Wun-hua Road, Huwei, Yunlin, Taiwan 632
E-mail: hyhuang@nfu.edu.tw

Weir-Sheng Shih

Wen-Hsing Hsu

National Tsing Hua University
Department of Electrical Engineering
101, Sec. 2, Kuang-fu Road, Hsinchu, Taiwan 300

Abstract. We present a framework to classify the film categories based on low-level features and visual effect features. This approach can serve as a prefilter that represents the movies that you want to watch from the Internet or movie-on-demand service. Our current domain of study is the movie preview. In our approach, we categorize films into three broad kinds: action, drama, and thriller films. Four low-level video features (average shot length, color variance, lighting key, and motion content) and visual effects are combined in our system to provide helpful information to demonstrate the film category. The results indicate that visual effect features are effective and useful information and play an important role for classifying the film categories. © 2008 SPIE and IS&T.
[DOI: 10.1117/1.2907215]

1 Introduction

What is a film? Film is a means of expression. Directors, actors, and cinematographers use this medium as a means to communicate a precisely crafted story line. This communication operates at several levels: explicitly, with the delivery of lines by the actors, and implicitly, with the background music, lighting, camera movements, and such. This fact in film making suggests that knowledge of cinematic principles can be exploited effectively for the understanding of films.¹ As the development of Internet and computer techniques makes it possible to watch a video via Internet, more and more people would like to watch a video at home or the office. Although we can watch or download movies from anywhere, these movies may not be suitable for every one. For example, children can download a movie and watch the violence or be scared by ghost, so parents need a system to protect their children. Hence, we propose a framework for genre classification based on low-level features and visual effect features from film previews.

Most approaches for shot boundary detection are to compute interframe distance from the decompressed video.^{2,3} Shot transitions can be detected by monitoring this

distance for significant changes. More common approaches are to use histograms of frame feature data. Approaches using global histogram represent each frame as a single vector, and those using localized histograms generate separating histograms for subsections of each frame. Interframe distances are calculated using simple vector-distance measures to compare corresponding histograms. Localized histograms, used in conjunction with additional features such as edge detection, perform well when applied in the TRECVID environment.^{3,4}

2 Proposed Method

To classify video type, first, we collected all movies that were played in Taiwan from 2004 to 2006.⁵ According to these statistical data, the action/adventure, drama (including comedy, drama, romance), and thriller (or horror) movies make up almost 88% of the movies in each year, which means if we can classify these three types of movies, then we can classify most movies.

2.1 Shot Boundary Detection

Shot boundary detection is used in the video processing scheme to estimate the relationship between frames. In our approach, we extend this algorithm proposed in Ref. 6 for the detection of shot boundaries using color histogram intersection for the hue-saturation-value space. Let $S(i)$ represent the intersection of histograms H_i and H_{i-1} of frames i and $i-1$, respectively. It is expressed as

$$S(i) = \sum_{j \in \text{allbins}} \min[H_i(j), H_{i-1}(j)]. \quad (1)$$

The magnitude $S(i)$ is often used as a measure of shot boundary in related works. The values of i where $S(i)$ is less than a fixed threshold are assumed to be the shot boundaries. To improve the accuracy, an iterative smoothing of the one-dimensional S function is performed first. We have adapted the algorithm reported in Ref. 7 based on the anisotropic diffusion. This is done in the context of

Paper 07182R received Sep. 4, 2007; revised manuscript received Nov. 9, 2007; accepted for publication Nov. 10, 2007; published online Apr. 17, 2008.

1017-9909/2008/17(2)/023011/5/\$25.00 © 2008 SPIE and IS&T.

scale space. S is smoothed iteratively using a Gaussian kernel such that the variance of the Gaussian function varies with the signal gradient. Formally,

$$S^{t+1}(i) = S^t(i) + \lambda[c_E \nabla_E S^t(i) + c_W \nabla_W S^t(i)], \quad (2)$$

where t is the iteration number and $0 < \lambda < 1/4$ with

$$\nabla_E S(i) \equiv S(i+1) - S(i), \quad (3)$$

$$\nabla_W S(i) \equiv S(i-1) - S(i). \quad (4)$$

The condition coefficients are a function of the gradients and are updated for every iteration

$$c_E^t = g(|\nabla_E S^t(i)|), \quad (5)$$

$$c_W^t = g(|\nabla_W S^t(i)|), \quad (6)$$

where $g(\nabla_E S) = \exp^{-(|\nabla_E|/k)^2}$ and $g(\nabla_W S) = \exp^{-(|\nabla_W|/k)^2}$. In our experiments, the constants were set to $\lambda=0.1$ and $k=0.1$. Finally, the shot boundaries are detected by finding the local minima in the smoothed similarity function S . For average shot length, it is directly computed by dividing the total number of frames by the total number of shots in the preview.

2.2 Low-Level Visual Effect Features

We found that human feeling can be affected via movie rhythm. To analyze the rhythm, we define the moving effect features of shot change, which are denoted *slow moving effect* and *fast moving effect* to understand the tempo property.⁸

2.2.1 Slow moving effect

This visual effect is a gradual change from shot to shot. Fade and dissolve are classified as this effect, because they usually need a long time duration in shot change.

2.2.2 Fast moving effect

This visual effect is two or more hard changes for the duration of a short time. This effect is mainly used in the classification of action and thriller films. These frames basically possess high contrast compared with their neighbor frames. In our experiments, if at least two abrupt cuts occurred in 0.2 s, we determined that the shot had a fast moving effect.

Abrupt cut. The total brightness difference will almost equal the brightness difference at the change frame, because there is only one rapid change in these durations.

Fade. Film minima brightness can be found in these durations because the fade must change between the original shot to black frames. Besides that, brightness change is gradually increased or decreased.

Dissolve. This technique is difficult to determine by brightness. Film minima brightness would not happen in these durations, because otherwise it becomes a fade. The brightness change can be gradually increased, decreased, or

stay unchanged. If a shot change can be detected and cannot be recognized as abrupt cut or fade, we classified this shot change as dissolve.

We only need to find fast and slow moving effects. Thus, we designed a rule to classify these two change modes (denoted as CH), which we present as abrupt cut and gradual change (including fade and dissolve). From the change mode, B denotes the neighboring frame's brightness, and B_d represents the brightness difference between two frames. The classification rule is defined as

$$\begin{cases} CH \in \text{abrupt cut}, & \text{if } 0.8 < T, \\ CH \in \text{gradual change}, & \text{otherwise,} \end{cases} \quad (7)$$

where $T = [\max(b) - \min(B)] / \max[\text{abs}(B_d)] < 1.2$. This threshold boundary is an empirical value for our testing data. With this classification rule and the visual effect definition, we can extract the fast and slow moving effects.

By using the frame numbers that were detected as fast or slow moving, we calculated the distance between two neighboring frames and obtained a new vector. Quantizing this vector, we compute the relative histograms normalized to 1. It can obtain two new features defined as fast moving effect distributions (f_v) and slow moving effect distributions (s_v). In our experiments, we set the total length of visual effect distribution as 100. We will further estimate the values of these two features denoted as F_{me} and S_{me} and expressed as

$$F_{me} = \sum_{l=1}^n l \times f_v(l), \quad S_{me} = \sum_{l=1}^n l \times s_v(l), \quad (8)$$

where l is the length index of visual effect distribution, and n is set to 100. If one kind of visual effect never occurred in a film, we set its visual effect value as 100. The visual effect value can serve as the expected value of how often this visual effect occurred in a film. That is, the lower the value, the more frequent the visual effect.

2.2.3 Lighting key

Lighting is an important dramatic agent and can also be used to direct the attention of the viewer to a certain area of importance in the scene. There are many ways to illuminate a scene. We adopt the scene quantity $\zeta_i(\mu, \sigma)$, which is defined as a measure of the lighting key of a frame $\zeta_i = \mu_i \times \sigma_i$, where μ and σ are the mean and standard deviation of i 'th key frame.^{1,9}

2.2.4 Motion content

The motion content represents the amount of activity in a film. Obviously, action films would have higher values for such a measure, and less visual disturbance would be expected for dramatic or romantic movies. To find visual disturbance, a method based on the structural tensor computation was used, which is described in Ref. 10.

2.3 Classification Processing

We chose the visual effect value and lighting key as primary features to build a classifier tree, as shown in Fig. 1.

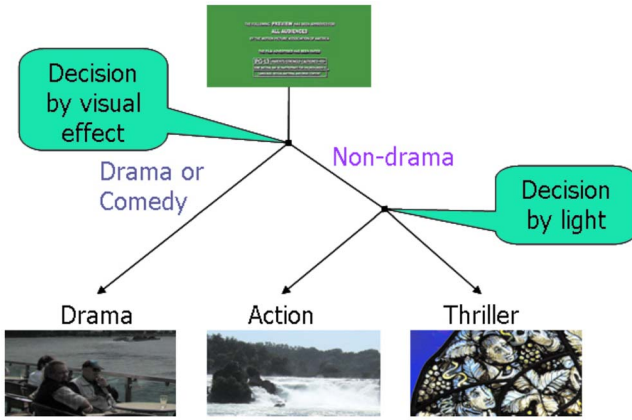


Fig. 1 Film classification method. A two-layer classifier and the corresponding decision rule.

Each branch of a decision tree represents a test outcome based on the decision rule, and the leaf nodes represent the classes or class distributions. The first decision rule is visual effect value, including the fast moving effect value F_{me} and the slow moving effect value S_{me} . We will calculate the threshold T_{fm} and T_{sm} to distinguish a film that belongs to the drama or nondrama category. For a film $F(i)$, the decision rule is defined as

$$\begin{cases} F(i) \in \text{nondrama}, & \text{if } F_{me} < T_{fm} \text{ and } S_{me} < T_{sm}, \\ F(i) \in \text{drama}, & \text{otherwise.} \end{cases} \quad (9)$$

The second decision rule is the lighting key. This feature can be used to distinguish action or thriller films. The lighting value is expressed as L . Classification trees calculate a threshold T_l and distinguish action and thriller films by a decision rule

$$\begin{cases} F(i) \in \text{action}, & \text{if } L > T_l, \\ F(i) \in \text{thriller}, & \text{if } L \leq T_l. \end{cases} \quad (10)$$

Thus, we can express all decision rules defined as

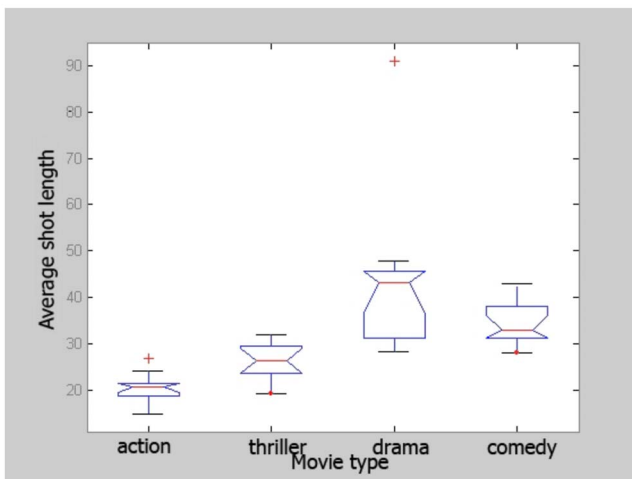


Fig. 2 Variability of average shot length. Boxes from left to right: action, thriller, drama, and comedy.

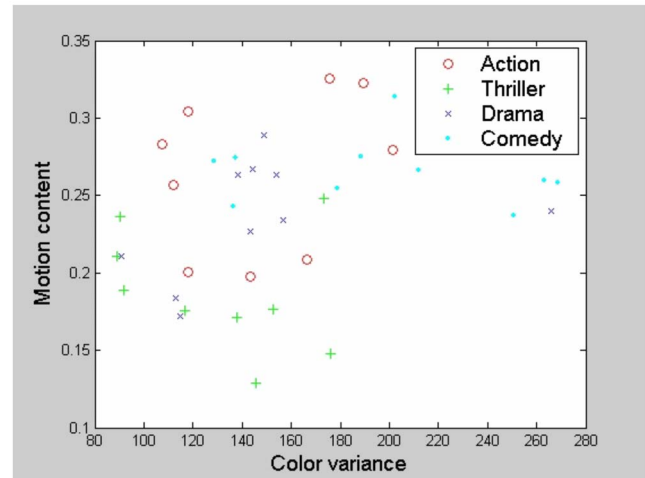


Fig. 3 Distribution of motion content and color variance.

$$\begin{cases} F(i) \in \text{action}, & \text{if } F_{me} < T_{fm} \text{ and } S_{me} < T_{sm} \text{ and } L > T_l, \\ F(i) \in \text{thriller}, & \text{if } F_{me} < T_{fm} \text{ and } S_{me} < T_{sm} \text{ and } L \leq T_l, \\ F(i) \in \text{drama}, & \text{otherwise.} \end{cases} \quad (11)$$

3 Experimental Results

We used 44 film previews to demonstrate our proposed method, including 9 thrillers, 10 actions, and 25 dramas (including comedies). For each preview, video tracks were analyzed at a frame rate of 12 f/s. We choose 6 film previews, and the total number of abrupt cuts and gradual changes (including fades and dissolves) is 484 and 176, respectively. Figure 2 shows the variability of average shot length. The box length is represented by the standard derivation of this cluster. From this figure, it is clearly known that the average shot length in action films is shorter than in drama and comedy films, which means the action film has a faster tempo than drama and comedy films do. Motion content and color variance distribution are shown in Fig. 3. From Fig. 3, the activity in an action film is higher than in

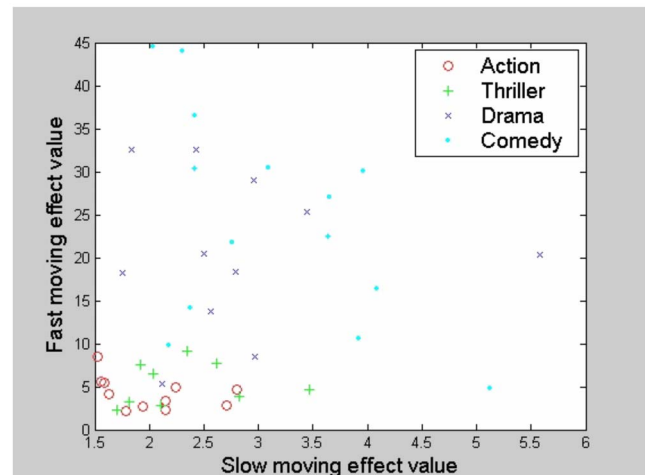


Fig. 4 Distribution of visual effect value.

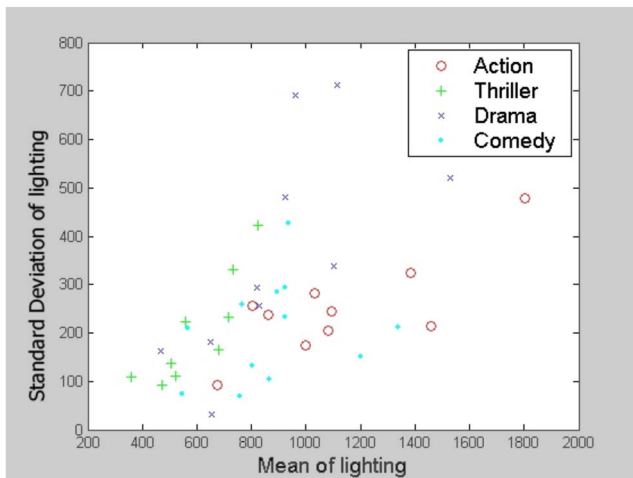


Fig. 5 Distribution of lighting key.

a drama. According to our experiments, it is difficult to distinguish film genres with color variance. Hence, these two features are not effective features to use to classify the film. Figure 4 shows the result of visual effect values. The action and thriller films tend to lower the fast moving effect value. The slow moving effect value has a low relationship with film genres. Figure 5 displays the distribution of the lighting key feature. From Fig. 5, it is clear that this feature cannot effectively categorize the film, hence, the lighting factor is not a critical feature.

The parameters of *Precision* and *Recall*¹¹ are used to evaluate the performance of visual effect detection. *Precision* and *Recall* are denoted the accuracy of detection and the ability of detection, respectively, expressed as

$$Recall = \frac{N_c}{N_a} \times 100\%, \quad Precision = \frac{N_c}{N_d} \times 100\%, \quad (12)$$

where N_c , N_a , and N_d are the number of correct detected, actual occurred, and detected, respectively. Table 1 presents the detection results of abrupt cut and gradual change. On the other hand, we indicate two feature sets F_a and F_b to evaluate the performance of these features. F_a contains the average shot length, lighting, and fast and slow moving effect values. F_b contains the average shot length, lighting, motion content, and color variance values. Table 2 presents the average precision of classification result corresponding to feature set. For this result, we demonstrate the visual effect features are more suitable than motion content and color variance.

Table 1 Detection results.

	N_c	N_a	N_d	Precision (%)	Recall (%)
Abrupt cut	467	484	486	96.09	96.49
Gradual change	173	176	232	74.57	98.01

Table 2 Average precision of classification result corresponding to feature set.

Feature Set	Action	Thriller	Drama	Total Average
F_a	85.19%	81.48%	74.07%	80.24%
F_b	62.96%	77.78%	40.74%	60.49%

4 Conclusions

In this paper, we have proposed a method to perform movie classification that combines four low-level features and visual effect features. The results indicate that combining visual cue with cinematic principles can provide powerful tools for genre categorization. In other words, our proposed approach can serve as a prefilter in selecting what kinds of movies you want to watch on the Internet.

Acknowledgments

The authors deeply appreciate the anonymous reviewers for their valuable suggestions. This work was supported in part by the National Science Council of Republic of China under Grants No. NSC 94-2213-E-007-055 and NSC 95-2221-E-150-091.

References

1. Z. Rasheed, Y. Sheikn, and M. Shah, "On the use of computable features for film classification," *IEEE Trans. Circuits Syst. Video Technol.* **15**(1), 52–63 (2005).
2. A. Hampapur, R. Jain, and T. Weymouth, "Digital video segmentation," in *Proc. 3rd ACM Int. Conf. Multimedia*, pp. 357–364 (1994).
3. T. Volkmer, S. M. M. Tahaghoghi, and H. E. Williams, "Gradual transition detection using average frame similarity," in *Conf. Comput. Vision Pattern Recogn. Workshop 9*, pp. 139–147 (2004).
4. A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVID," in *Proc. 8th ACM SIGMM Int. Workshop on Multimedia Inf. Retrieval*, pp. 321–330 (2003).
5. Available: <http://www.truemovie.com/>
6. G. Rafae and E. Richard, *Digital Image Processing*, Prentice Hall, Upper Saddle River, N.J. (2002).
7. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(7), 629–639 (1990).
8. H. Y. Huang, W. S. Shih, and W. H. Hsu, "A movie classifier based on visual features," in *Proc. 12th Int. Conf. Comput. Anal. Image Patterns*, W. G. Kropatsch, M. Kampel, and A. Hanbury, Eds., Lect. Notes Comput. Sci. 4673, 937–944 (2007).
9. Pacific Cinematheque, "Lighting in filmmaking." Available: <http://www.inpoint.org/>
10. B. Jahne, *Spatio-Temporal Image Processing: Theory and Scientific Applications*, Springer-Verlag, Berlin (1991).
11. G. Lupatini, C. Saraceno, and R. Leonardi, "Scene break detection: a comparison," in *Proc. 8th Int. Workshop on Research Issues in Data Engineering, Continuous-Media Databases and Applications*, A. Silberschatz, A. Zhang, and S. Mehrotra, Eds., pp. 34–41, IEEE Computer Society Press, Los Alamitos, CA (1998).



Hui-Yu Huang received her BS in electronic engineering from Feng Chia University, Taiwan, in 1992; her MS in electrical and computer engineering from Yuan-Ze Institute of Technology, in 1994, and her PhD in electrical engineering from National Tsing Hua University, Taiwan, in 2002. Since 2005, she has been with the Department of Computer Science and Information Engineering at National Formosa University in Taiwan, where she is now an assistant professor. Her research interests include multimedia processing, neural networks, pattern recognition, and content-based image/video retrieval. Dr. Huang is a member of the Institute of Electrical

and Electronics Engineers and the Chinese Association of Image Processing and Pattern Recognition.



Weir-Sheng Shih received his BS in aeronautical engineering from National Formosa University, Taiwan, in 2003, and his MS degree in electrical engineering from National Tsing Hua University, Taiwan, in 2005. His research interests include image/video coding and multimedia systems.



Wen-Hsing Hsu received his BS in electrical engineering from National Cheng Kung University, Taiwan, in 1972, and his ME and PhD in electrical engineering from Keio University, Japan, in 1978 and 1982, respectively. In 1982, he joined in the Department of Electrical Engineering at National Tsing Hua University in Taiwan, where he is now a professor. His research interests include image processing, biologic identification, and network security.

Dr. Hsu is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan, and the Chinese Association of Image Processing and Pattern Recognition.